Taxonomies & Data

Not Just For

Classification

E-Records Conference 2012

COMPU-DATA
International, LLC

Preserving

the

World's Knowledge

Available Anytime Anywhere℠

COMPU-DATA International, LLC

Company Overview

Who are we?

CDI is a leading information management integrator based in Spring, Texas (North of Houston) with offices in Dallas, Miami, FL and the Washington, D.C. area. We have been in business for over 24 years with 18 of those focused in Content and Data Integration (CADI™), enterprise search, classification, capture and data management. We are a small business and designated a certified Texas HUB contractor.

What do we do?

Integration, software development and reseller of best-of-breed products for ECM solutions focused in Search, Automatic Classification, Capture and Business Automation (Workflows). We work with Government and private industry customers in delivering successful departmental and enterprise solutions.

Who do we serve?

Medium to large organizations in government, health care, manufacturing and oil industries.

RMDA Records Management and Declassification Agency

POLICE ANCHORAGE ALASKA

Texas State Library and Archives Commission

GEISINGER HEALTHCARE PROFESSIONALS

MICHELIN A better way forward

KBR

Honeywell

COMPU-DATA International, LLC

Preserving the Worlds Knowledge - Available Anytime Anywhere℠

Juan J. Celaya, President & CEO

COMPU-DATA International, LLC

jcelaya@cdlac.com

www.cdlac.com **blog.cdlac.com**

Microsoft Partner
Silver Independent Software Vendor (ISV)

Microsoft
Authorized Education Reseller

The experts in the Capture and Enhancement of Data for storage, retrieval and collaboration purposes.

DIR

Contract #s
DIR-SDD-1769
DIR-SDD-1728

CCG

Contract #
CCG-DIS-2010-003

STATE OF TEXAS Certified HUB

COMPU-DATA International, LLC

Preserving the Worlds Knowledge - Available Anytime Anywhere[SM]

Presentation Overview

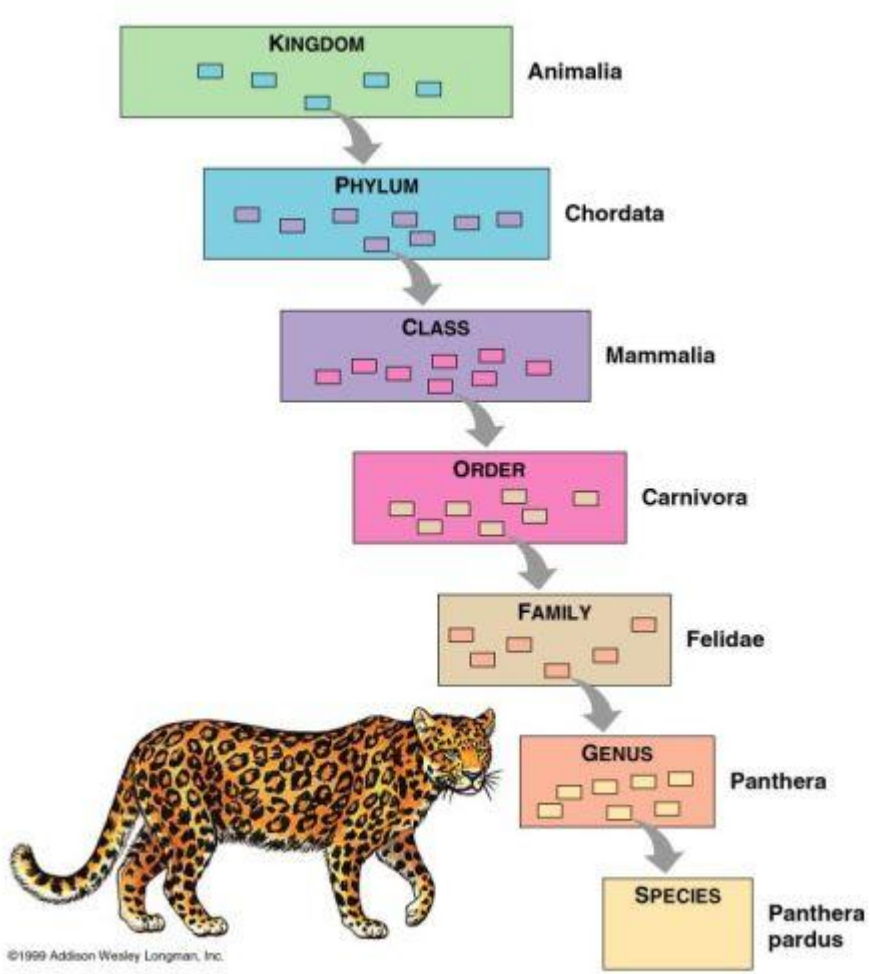
- Using Taxonomies
  - For Information Management
  - Not just for Classification
  - Solution Results
- Initial Setup Sample
  - Rules
  - Taxonomy
- Leveraging Automatically Applied Metadata to Deliver:
  - Fully Automated Document Library Permissions
  - Information Management Policy Settings
  - Information Rights Management.
- Example of utilizing the combined power of:
  - SharePoint 2010
  - Data Enhancement System
  - conceptClassifier for SharePoint
  - Content Type Updater

COMPU-DATA International, LLC

What is a Taxonomy?

In Information Management!

NOT in its traditional use for the classification of plants and animals.



COMPU-DATA International, LLC

Taxonomy means:
The classification of something.

Classification means:
The action or process of grouping or creating a set of something according to shared qualities or characteristics.

For today, Taxonomy in Information Management means:

The hierarchical arrangement into groups, classes, categories or facets of things based on their shared qualities, properties or other established criteria such as content.

Most common use for a taxonomy in Information Management is to classify, categorize and organize content.

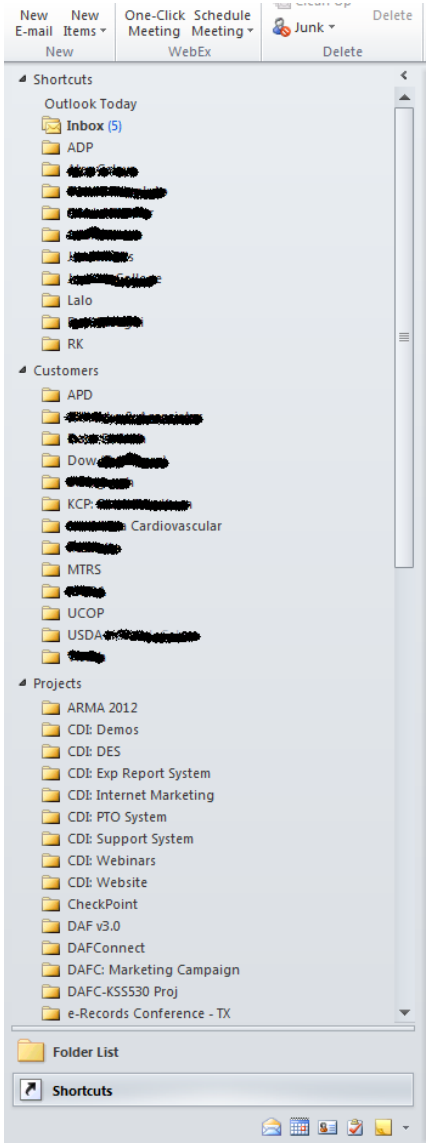
Who is currently Using a Taxonomy?

COMPU-DATA International, LLC

We all are!

PRIMARILY USED TO IMPROVE SEARCH RESULTS.

Preserving the Worlds Knowledge - Available Anytime Anywhere[SM]



My Outlook Inbox:

- 90+ Shortcuts
- 3GB of content
- A few hundred folders used for manual classification

Taxonomies & Data
Not just for Classification

Taxonomies and auto-tagging engines can be used in a much broader context to bring real value to your organization!

Just as we use Taxonomies to classify content for search we can use Taxonomies to automatically tag the content:

1. Improve Search Relevance by using our own Vocabularies.
2. Enforce Records Management Policies.
3. Compliance with State and Federal Regulations.
    a. Texas Public Information Act
    b. Federal Freedom of Information Act (FOIA)
    c. Health Insurance Portability & Accountability Act (HIPAA)
4. Information & General Governance Standards.
5. Enterprise Metadata Management.
6. Sensitive Information Identification & Protection.
    a. Personally Identifiable Information (PII)
    b. Protected Health Information (PHI)
    c. Exceptions to FOIA and Texas Public Information Act
7. Data Migration.

Solution Results

Taxonomies and auto-tagging engines can be used in a much broader context to bring real value to your organization!

Automatic Tagging of Content allows:

Process Automation
driven by how content is tagged!

Now we can:

1. Automatically enforce Records and Information Rights Management.
2. Automatic Content & File Type application to Content.
3. Automatic Migration of Content to Document Libraries.
4. Find, Store, Preserve, Secure & Control Data Assets Across Distinct Business & Service Delivery Units.
5. Ensures Metadata Tagging Consistency & Data Transparency at Every Level.
6. Promote “Secure Collaboration” Through Time-Saving Automation of Records Management & Information Assurance Activities.

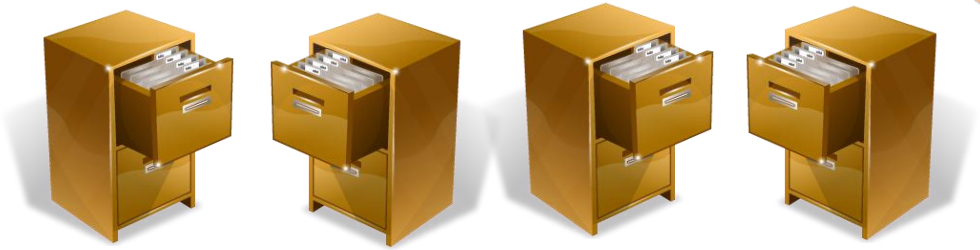
Real-Time Data Transparency
Records Declaration
Information Security

COMPU-DATA International, LLC

Manual Metadata Application

Is this a Record?
Is this a Sensitive Document?
Where do I put it?

Server Content with Appropriate Metadata, Retention Codes and Rights Management Templates



Document Library 1

Document Library 2

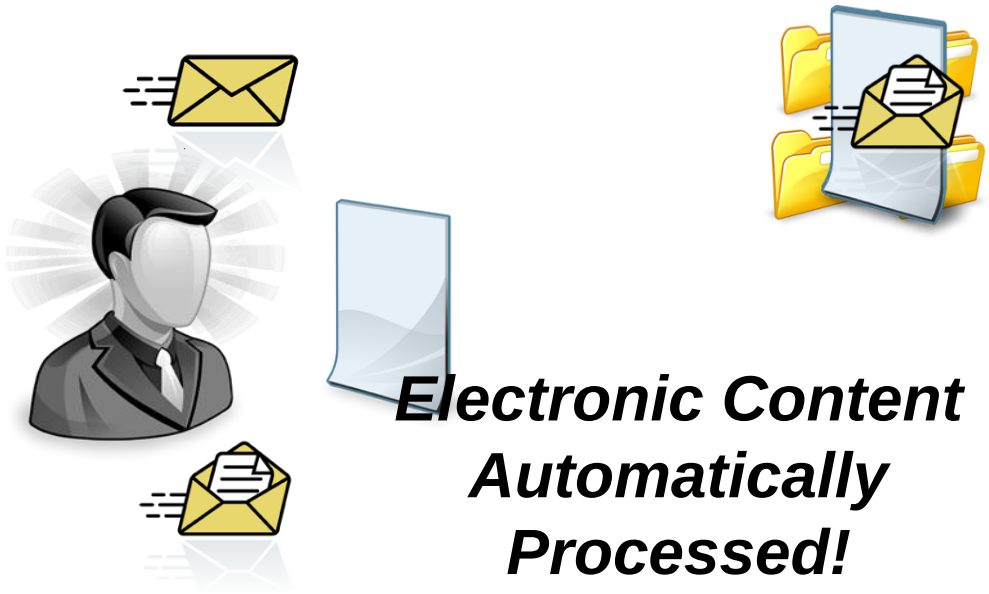
Document Library 3

Document Library 4

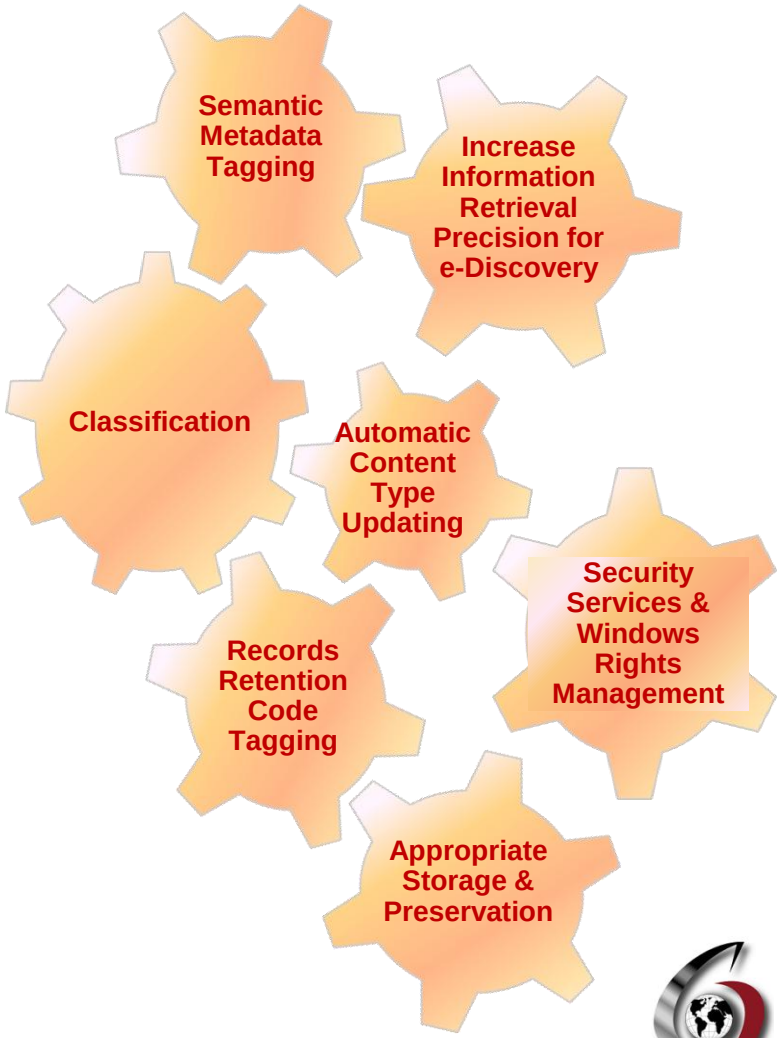
Preserving the Worlds Knowledge - Available Anytime Anywhere[SM]

COMPU-DATA International, LLC

Typical Scenario of Solution



Electronic Content Automatically Processed!

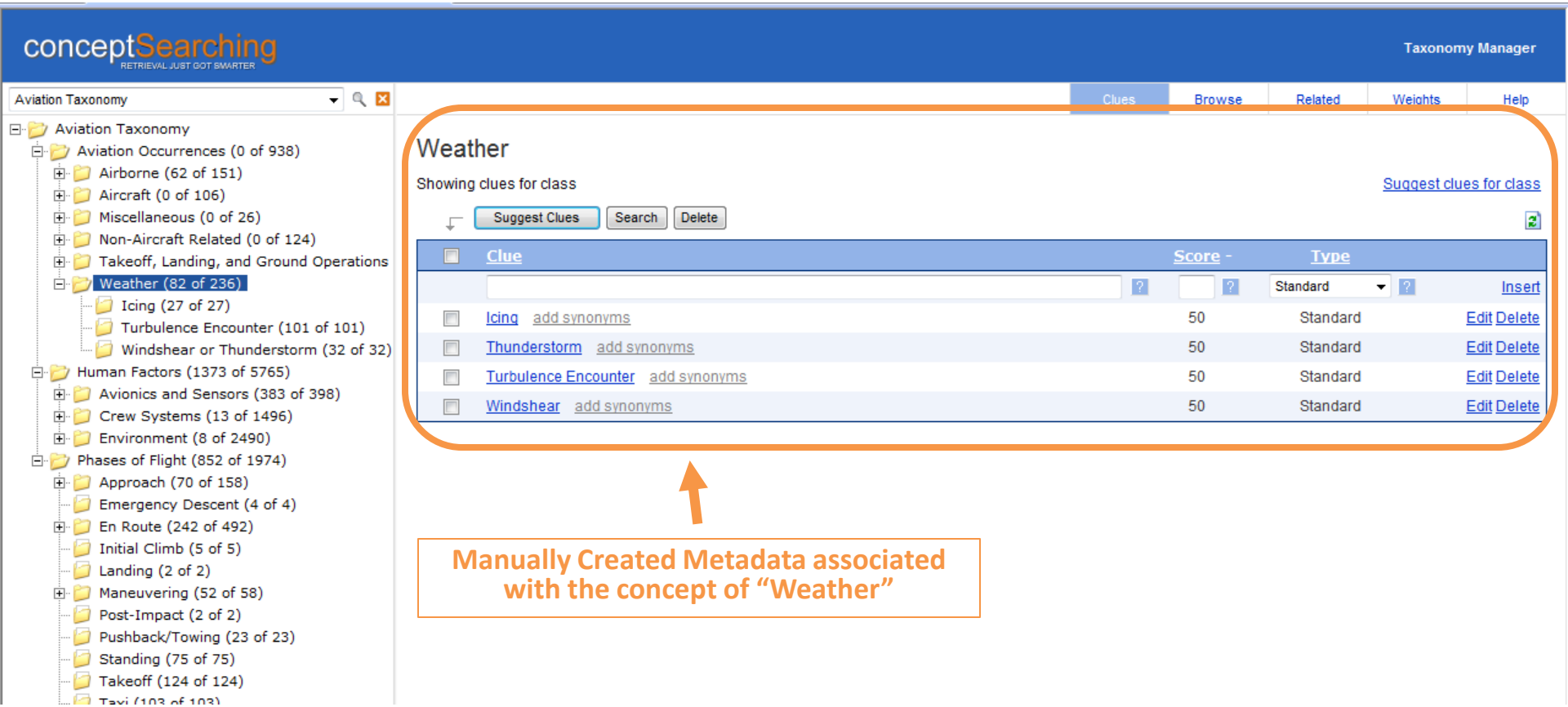


Preserving the Worlds Knowledge - Available Anytime Anywhere[SM]

Taxonomy Management

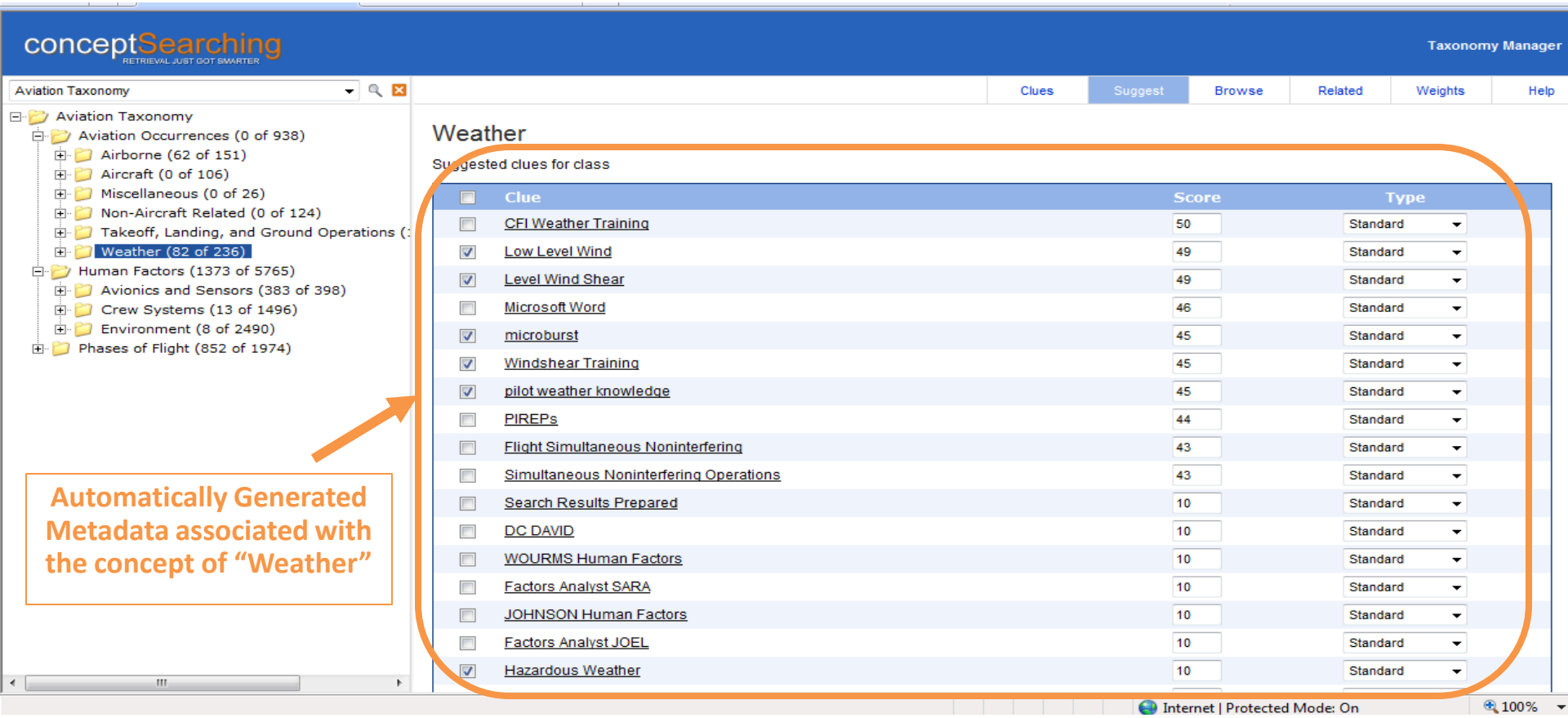
Enabling the Automatic Meta-tagging and Auto-Classification of Documents and Records

Each node is a piece of metadata that gets tagged to a document or record based upon the prevalence of a clue within the document



Automatic Metadata Generation

Unique IP of Compound Term Processing enables the identification of compound terms (not keywords) from highly relevant content that can be used to trigger the automatic meta-tagging and auto-classification processes

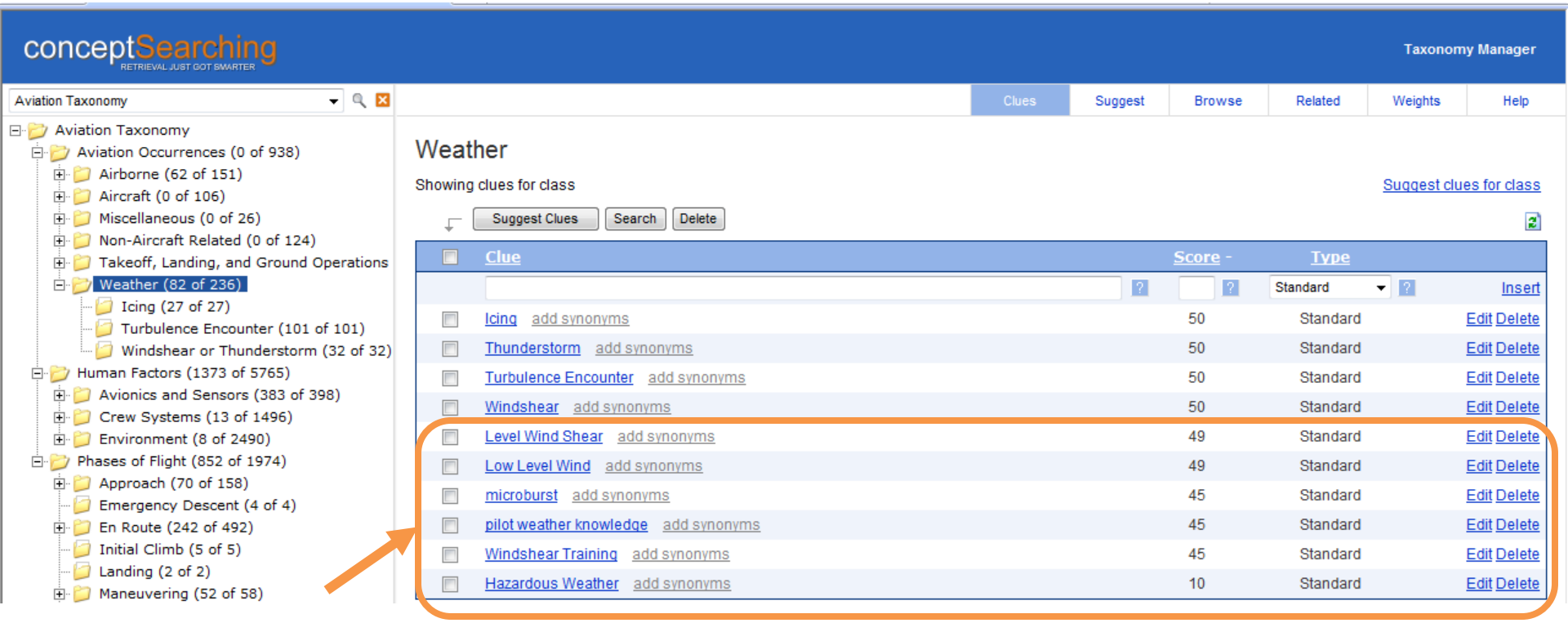


Automatically Generated Metadata associated with the concept of "Weather"

Automatic Metadata Generation

Automatically generated metadata is added to original metadata for the category/folder

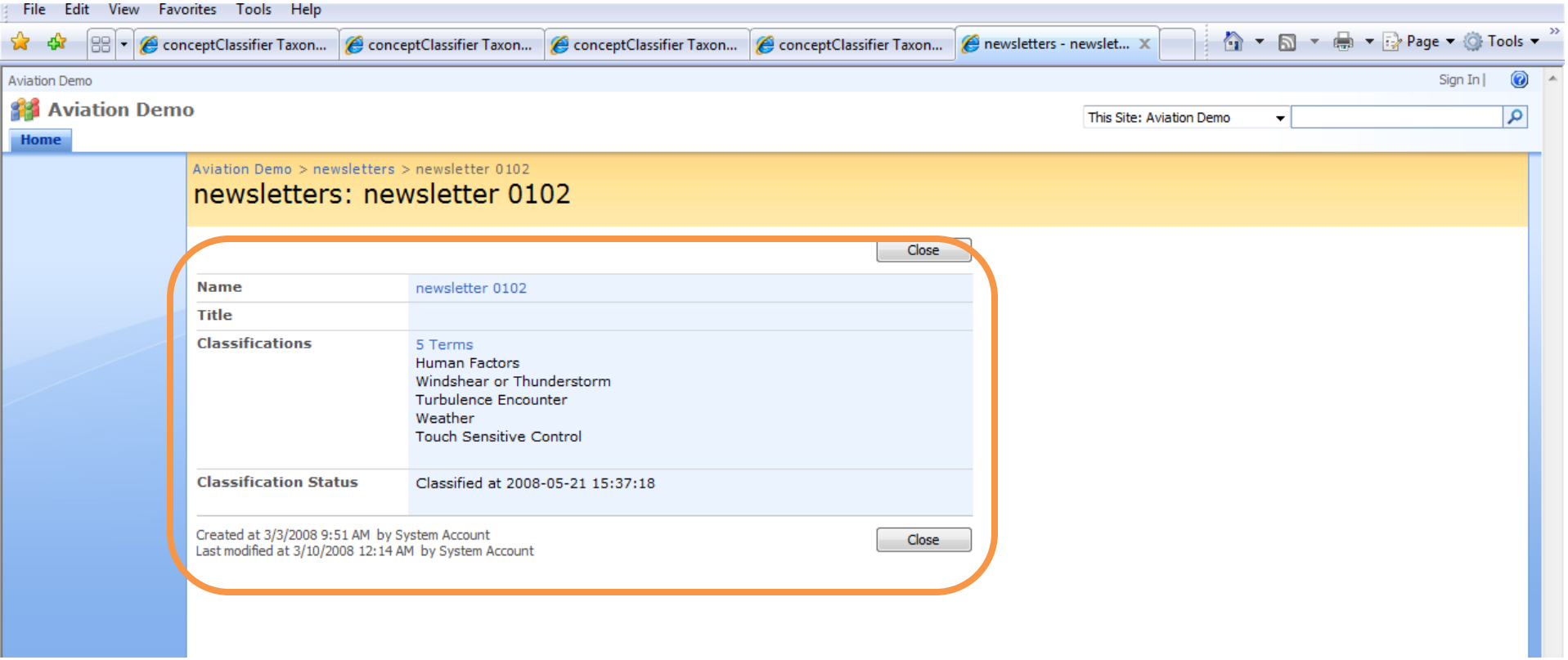
Outcome: more semantics that can be linked to a document or record result in information that becomes more actionable (the document/record is now retrievable and classifiable)



Highly relevant metadata generated by Taxonomy Manager added to original clue set for the concept of "Weather"

Automatic Meta-tagging

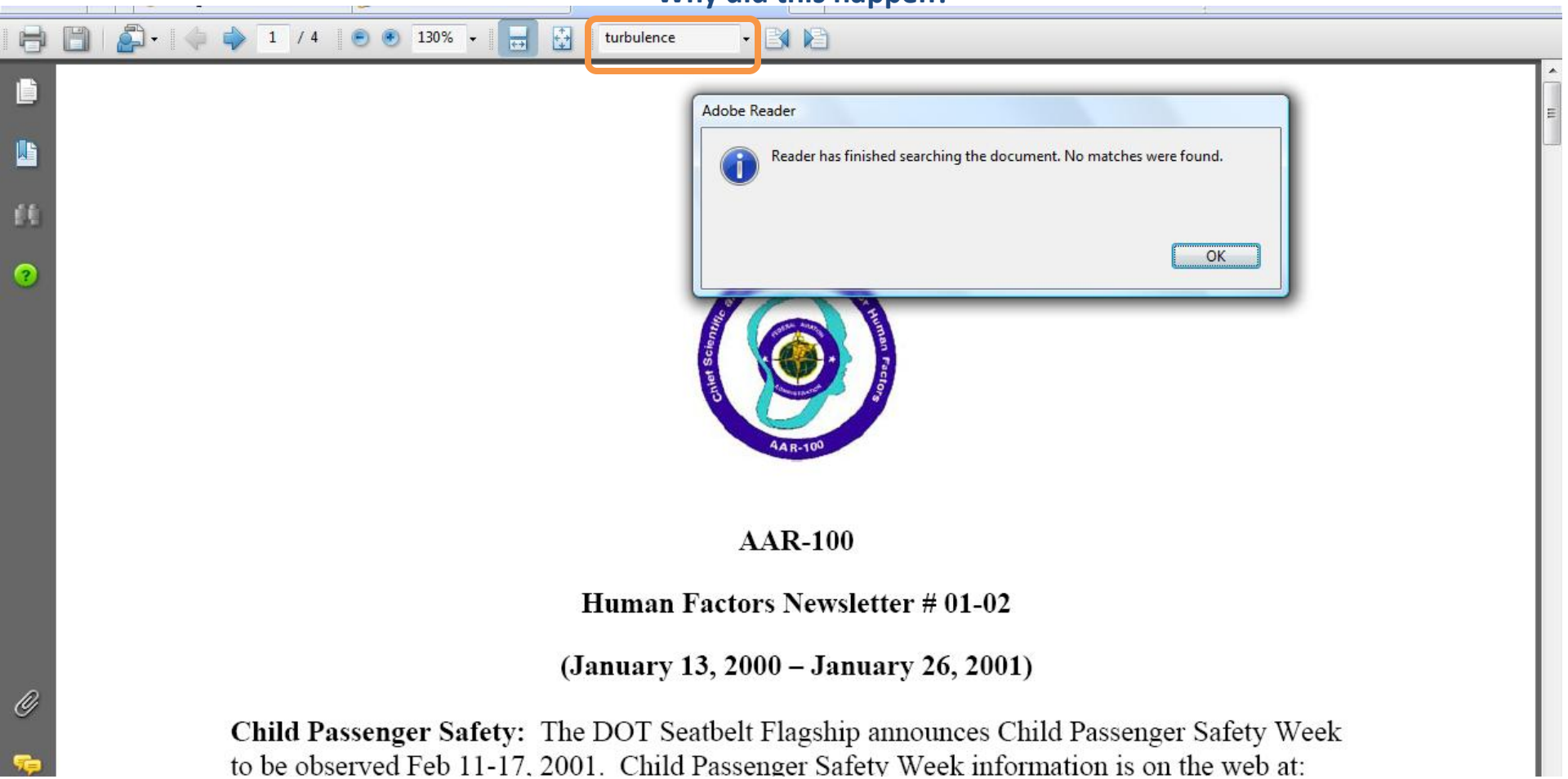
Metatags are automatically added to the properties field of each document making the document more valuable to the organization by increasing the ability of the document to be retrieved using enterprise search solutions that use keywords and metadata to retrieve information



Automatic Meta-tagging in Action

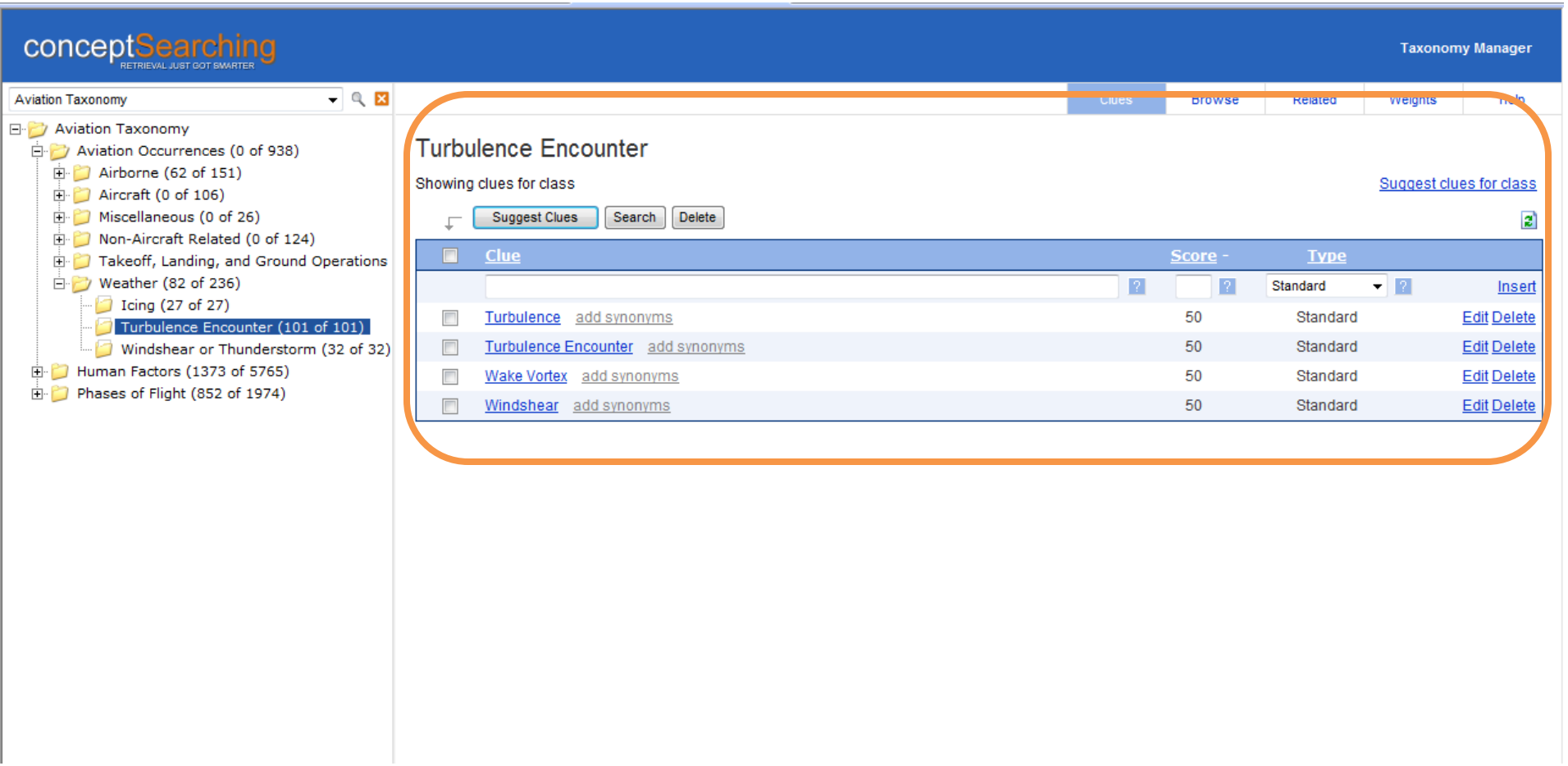
One of our Metatags for the Newsletter 01-02 was "Turbulence Encounter" however when we search for this term within the document we do not find it

Why did this happen?



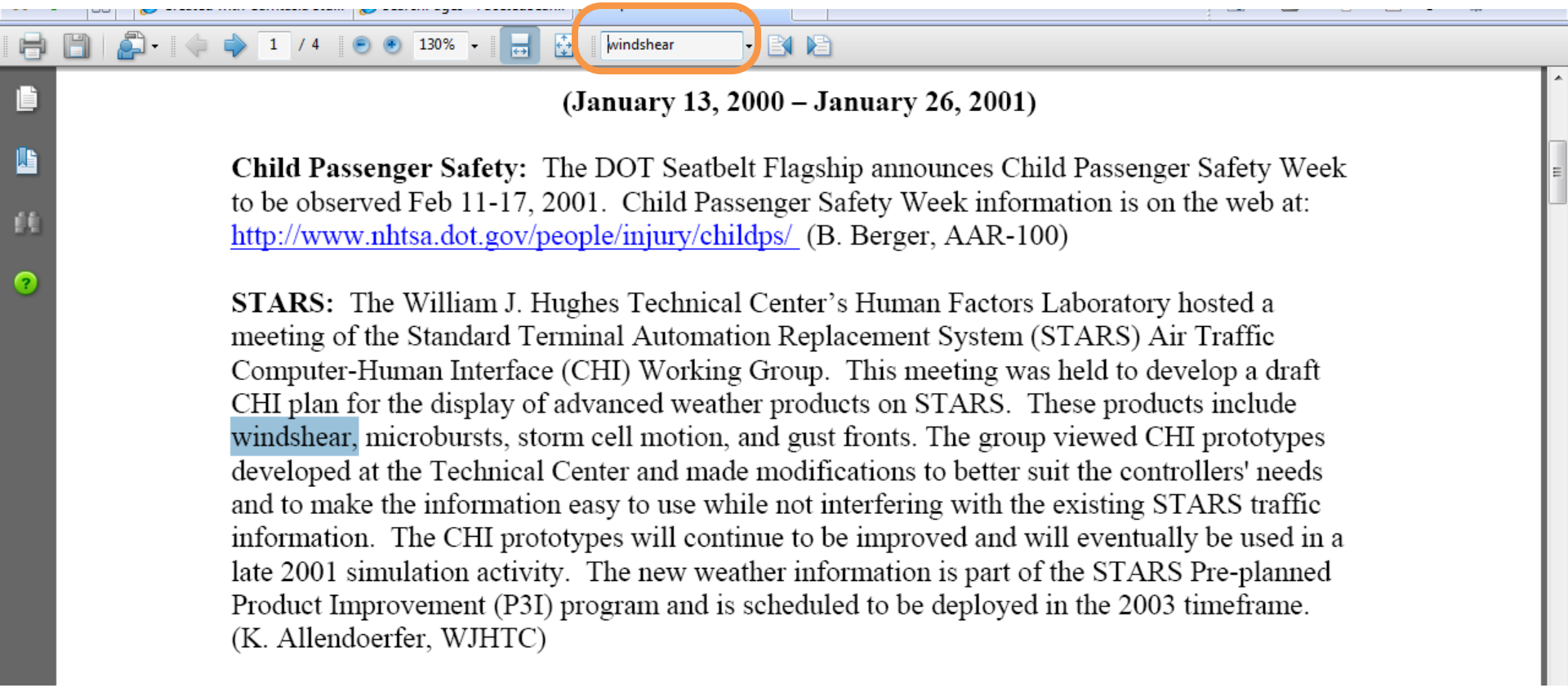
Automatic Meta-tagging in Action

Turbulence Encounter is only one of 4 "clues" that must exist within a document in order for that document to be automatically meta-tagged with the concept of Turbulence Encounter



Automatic Meta-tagging in Action

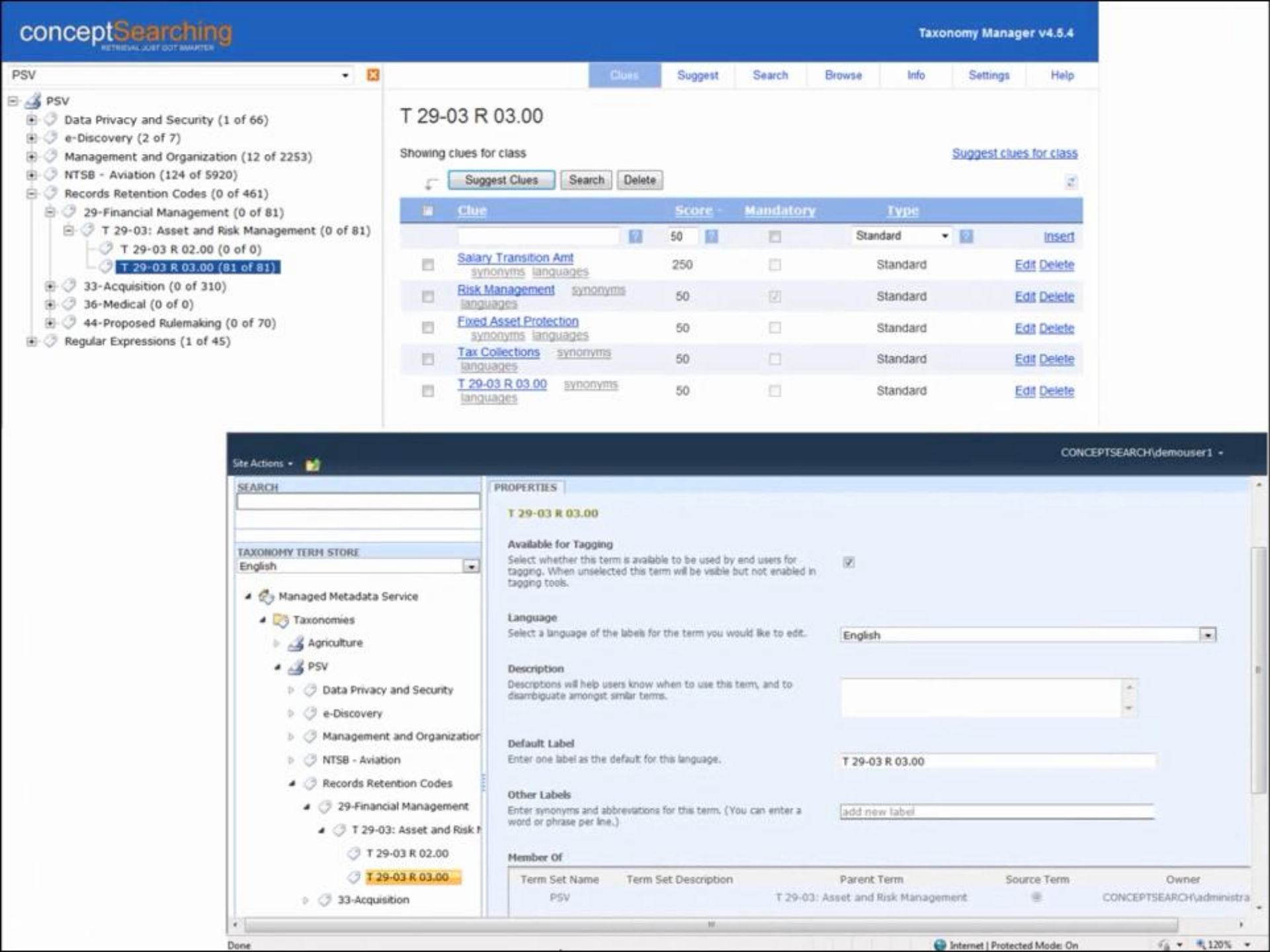
When we search our document using another clue for Turbulence Encounter, "Windshear", we see that its existence within the document triggered the automated meta-tagging event that resulted in the document being tagged with "Turbulence Encounter"

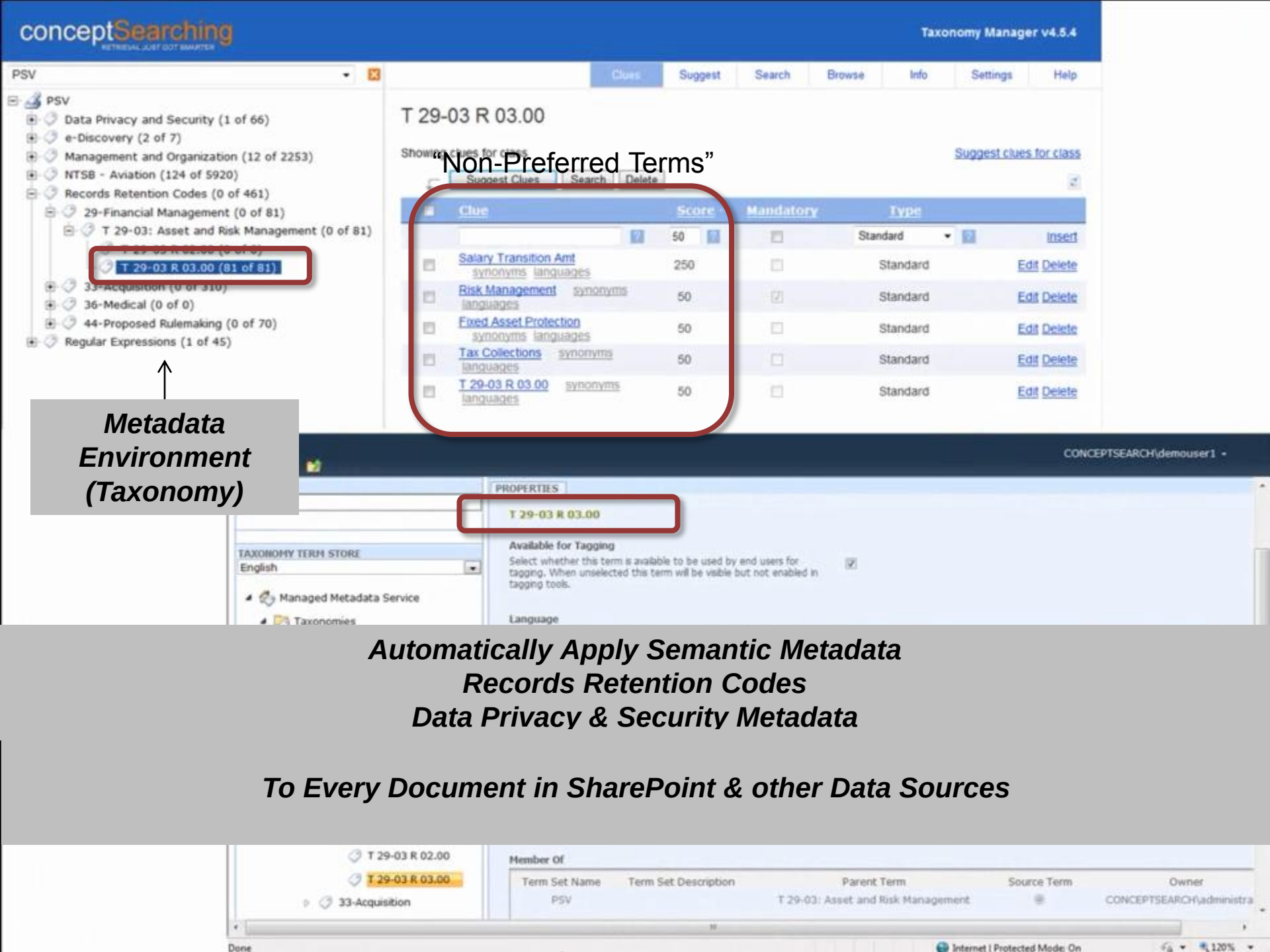


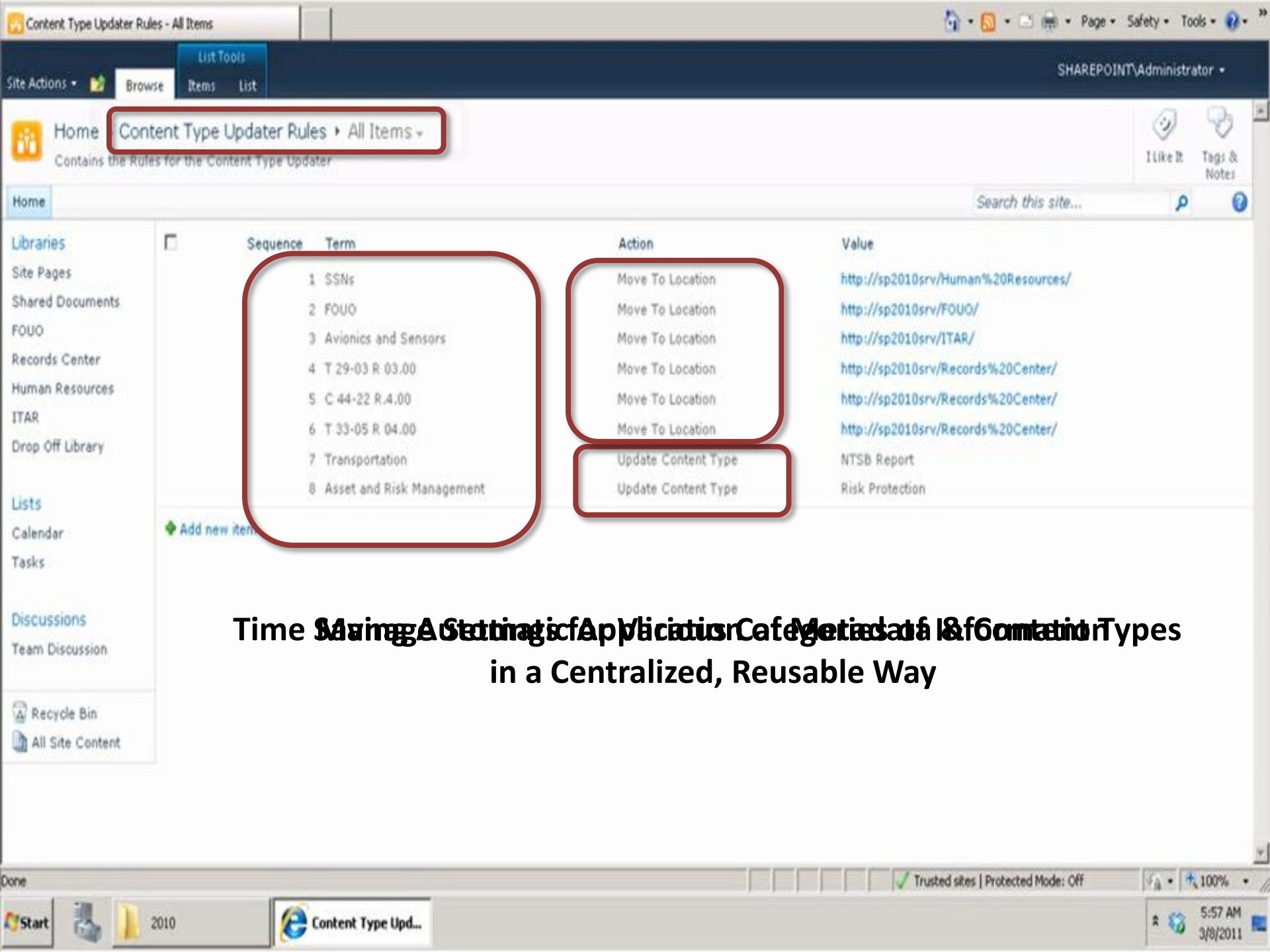
(January 13, 2000 – January 26, 2001)

Child Passenger Safety: The DOT Seatbelt Flagship announces Child Passenger Safety Week to be observed Feb 11-17, 2001. Child Passenger Safety Week information is on the web at: http://www.nhtsa.dot.gov/people/injury/childps/ (B. Berger, AAR-100)

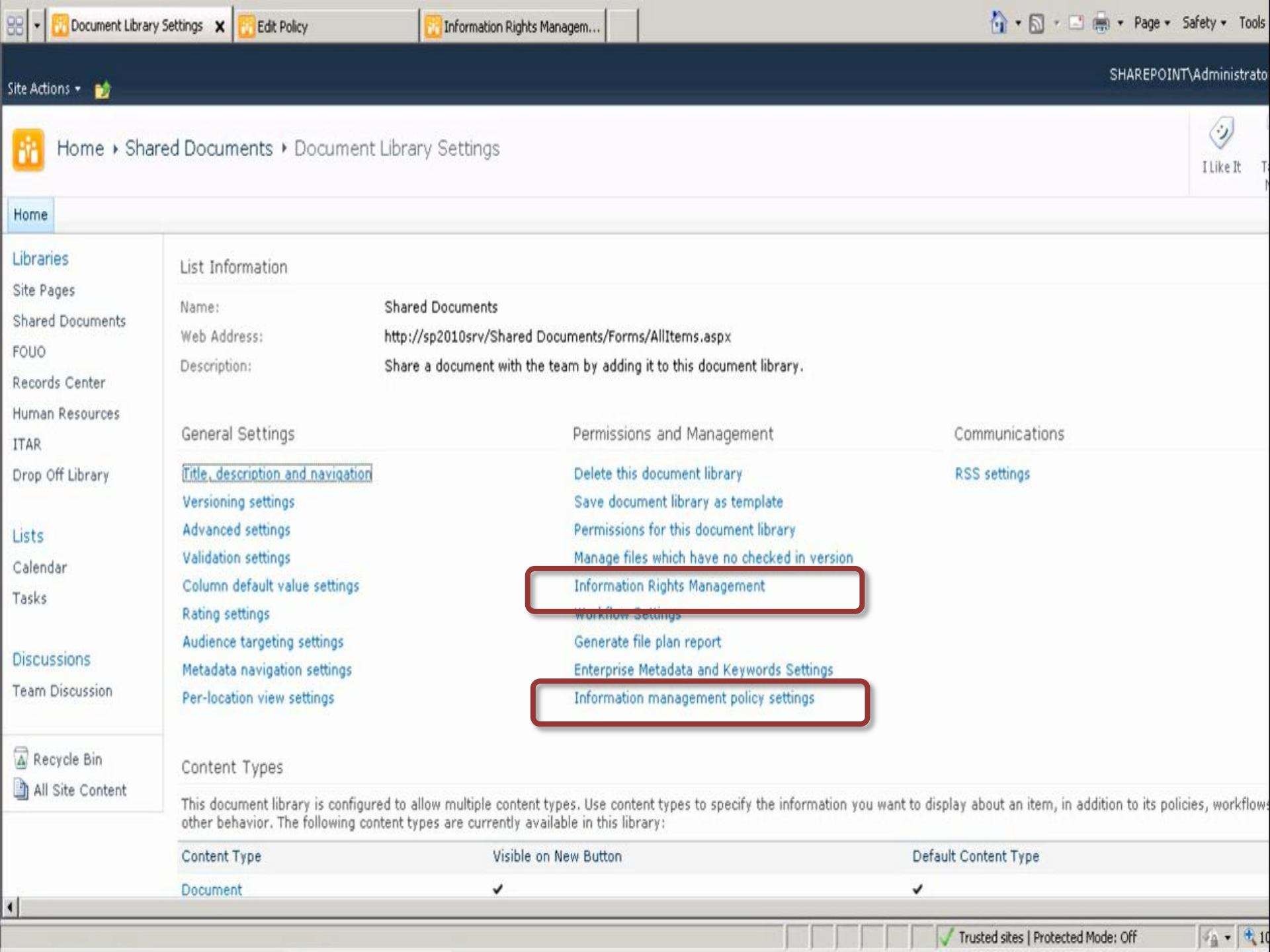
STARS: The William J. Hughes Technical Center's Human Factors Laboratory hosted a meeting of the Standard Terminal Automation Replacement System (STARS) Air Traffic Computer-Human Interface (CHI) Working Group. This meeting was held to develop a draft CHI plan for the display of advanced weather products on STARS. These products include windshear, microbursts, storm cell motion, and gust fronts. The group viewed CHI prototypes developed at the Technical Center and made modifications to better suit the controllers' needs and to make the information easy to use while not interfering with the existing STARS traffic information. The CHI prototypes will continue to be improved and will eventually be used in a late 2001 simulation activity. The new weather information is part of the STARS Pre-planned Product Improvement (P3I) program and is scheduled to be deployed in the 2003 timeframe. (K. Allendoerfer, WJHTC)

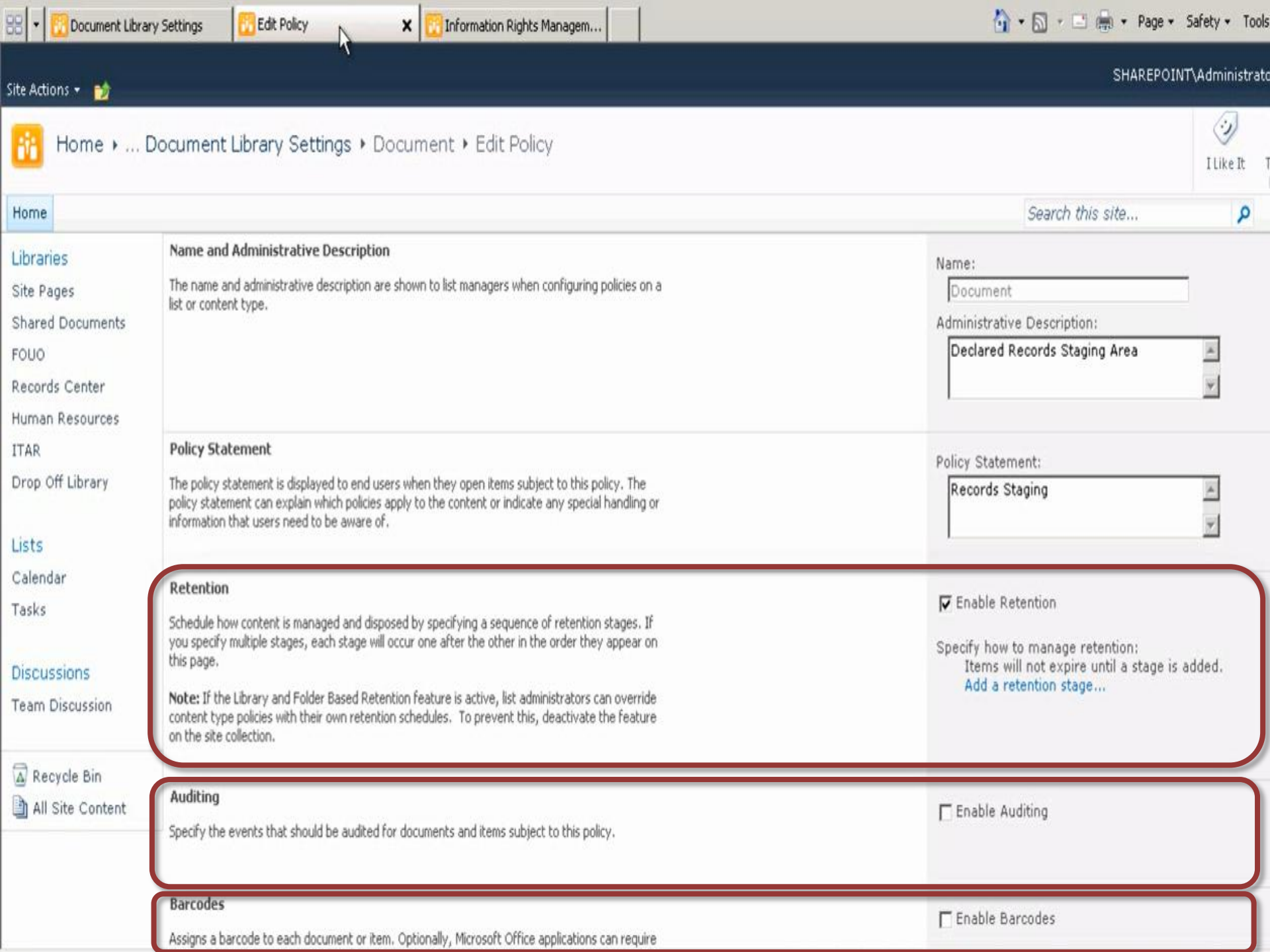

conceptSearching
Taxonomy Manager v4.5.4
PSV
Clues
Suggest
Search
Browse
Info
Settings
Help
PSV
Data Privacy and Security (1 of 66)
e-Discovery (2 of 7)
Management and Organization (12 of 2253)
NTSB - Aviation (124 of 5920)
Records Retention Codes (0 of 461)
29-Financial Management (0 of 81)
T 29-03: Asset and Risk Management (0 of 81)
T 29-03 R 02.00 (0 of 0)
T 29-03 R 03.00 (81 of 81)
33-Acquisition (0 of 310)
36-Medical (0 of 0)
44-Proposed Rulemaking (0 of 70)
Regular Expressions (1 of 45)
T 29-03 R 03.00
Showing clues for class
Suggest clues for class
Suggest Clues
Search
Delete
Clue
Score
Mandatory
Type
50
Standard
Insert
Salary Transition Amt
synonyms languages
250
Standard
Edit Delete
Risk Management
synonyms languages
50
Standard
Edit Delete
Fixed Asset Protection
synonyms languages
50
Standard
Edit Delete
Tax Collections
synonyms languages
50
Standard
Edit Delete
T 29-03 R 03.00
synonyms languages
50
Standard
Edit Delete
Site Actions
CONCEPTSEARCH\demouser1
SEARCH
TAXONOMY TERM STORE
English
Managed Metadata Service
Taxonomies
Agriculture
PSV
Data Privacy and Security
e-Discovery
Management and Organization
NTSB - Aviation
Records Retention Codes
29-Financial Management
T 29-03: Asset and Risk
T 29-03 R 02.00
T 29-03 R 03.00
33-Acquisition
PROPERTIES
T 29-03 R 03.00
Available for Tagging
Select whether this term is available to be used by end users for tagging. When unselected this term will be visible but not enabled in tagging tools.
Language
Select a language of the labels for the term you would like to edit.
English
Description
Descriptions will help users know when to use this term, and to disambiguate amongst similar terms.
Default Label
Enter one label as the default for this language.
T 29-03 R 03.00
Other Labels
Enter synonyms and abbreviations for this term. (You can enter a word or phrase per line.)
add new label
Member Of
Term Set Name
Term Set Description
Parent Term
Source Term
Owner
PSV
T 29-03: Asset and Risk Management
CONCEPTSEARCH\administra
Done
Internet | Protected Mode: On
120%

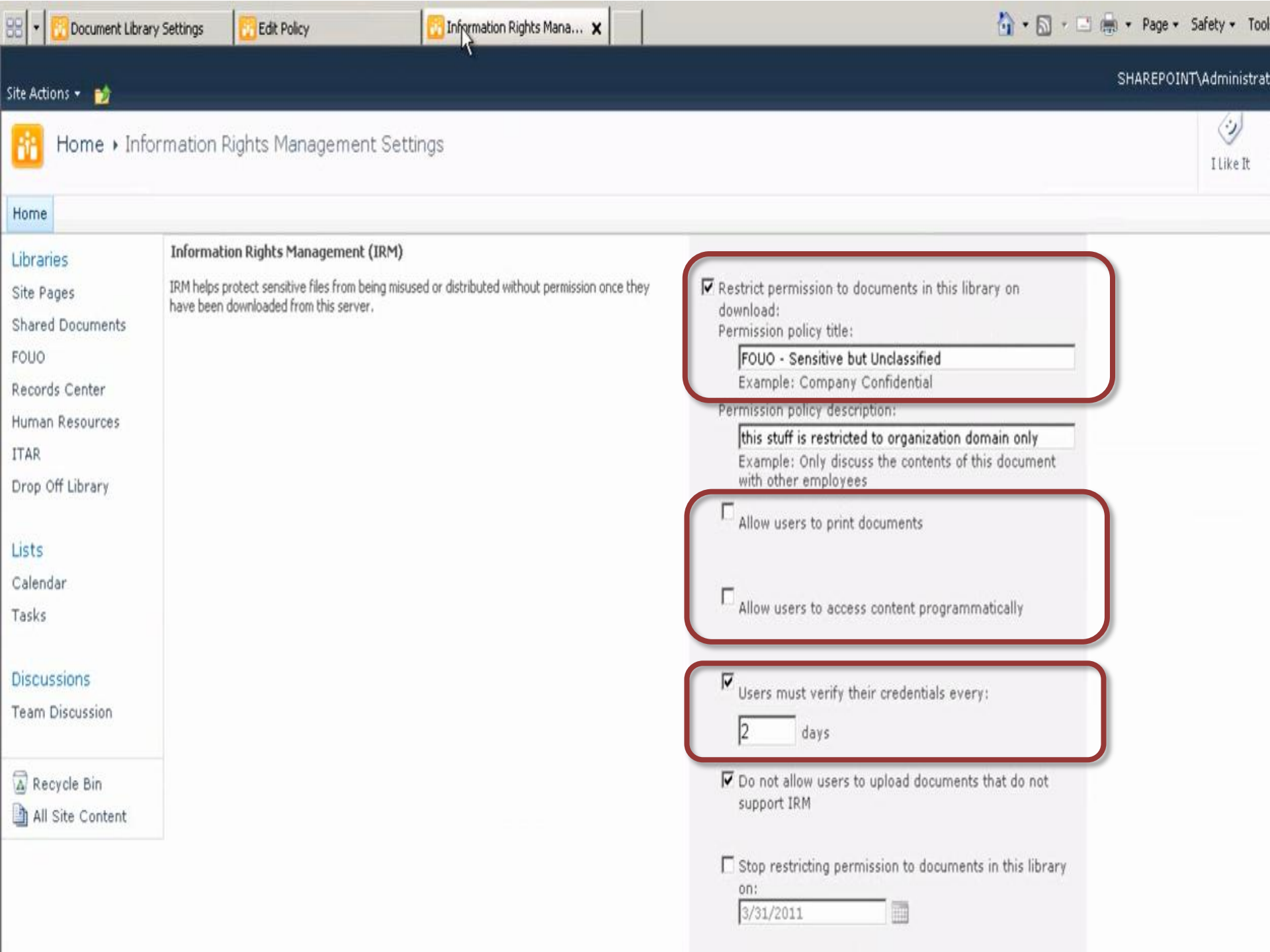

conceptSearching
Taxonomy Manager v4.5.4
PSV
Clues
Suggest
Search
Browse
Info
Settings
Help
Data Privacy and Security (1 of 66)
e-Discovery (2 of 7)
Management and Organization (12 of 2253)
NTSB - Aviation (124 of 5920)
Records Retention Codes (0 of 461)
29-Financial Management (0 of 81)
T 29-03: Asset and Risk Management (0 of 81)
T 29-03 R 03.00 (81 of 81)
36-Medical (0 of 0)
44-Proposed Rulemaking (0 of 70)
Regular Expressions (1 of 45)
T 29-03 R 03.00
Suggest clues for class
Suggest Clues
Search
Delete
Clue
Score
Mandatory
Type
Insert
Salary Transition Amt
synonyms languages
250
Standard
Edit Delete
Risk Management
synonyms languages
50
Standard
Edit Delete
Fixed Asset Protection
synonyms languages
50
Standard
Edit Delete
Tax Collections
synonyms languages
50
Standard
Edit Delete
T 29-03 R 03.00
synonyms languages
50
Standard
Edit Delete
"Non-Preferred Terms"
Metadata Environment (Taxonomy)
CONCEPTSEARCH\demouser1
PROPERTIES
T 29-03 R 03.00
TAXONOMY TERM STORE
English
Managed Metadata Service
Taxonomies
Available for Tagging
Select whether this term is available to be used by end users for tagging. When unselected this term will be visible but not enabled in tagging tools.
Language
Automatically Apply Semantic Metadata
Records Retention Codes
Data Privacy & Security Metadata
To Every Document in SharePoint & other Data Sources
T 29-03 R 02.00
T 29-03 R 03.00
33-Acquisition
Member Of
Term Set Name
Term Set Description
Parent Term
Source Term
Owner
PSV
T 29-03: Asset and Risk Management
CONCEPTSEARCH\administra
Done
Internet | Protected Mode: On
120%



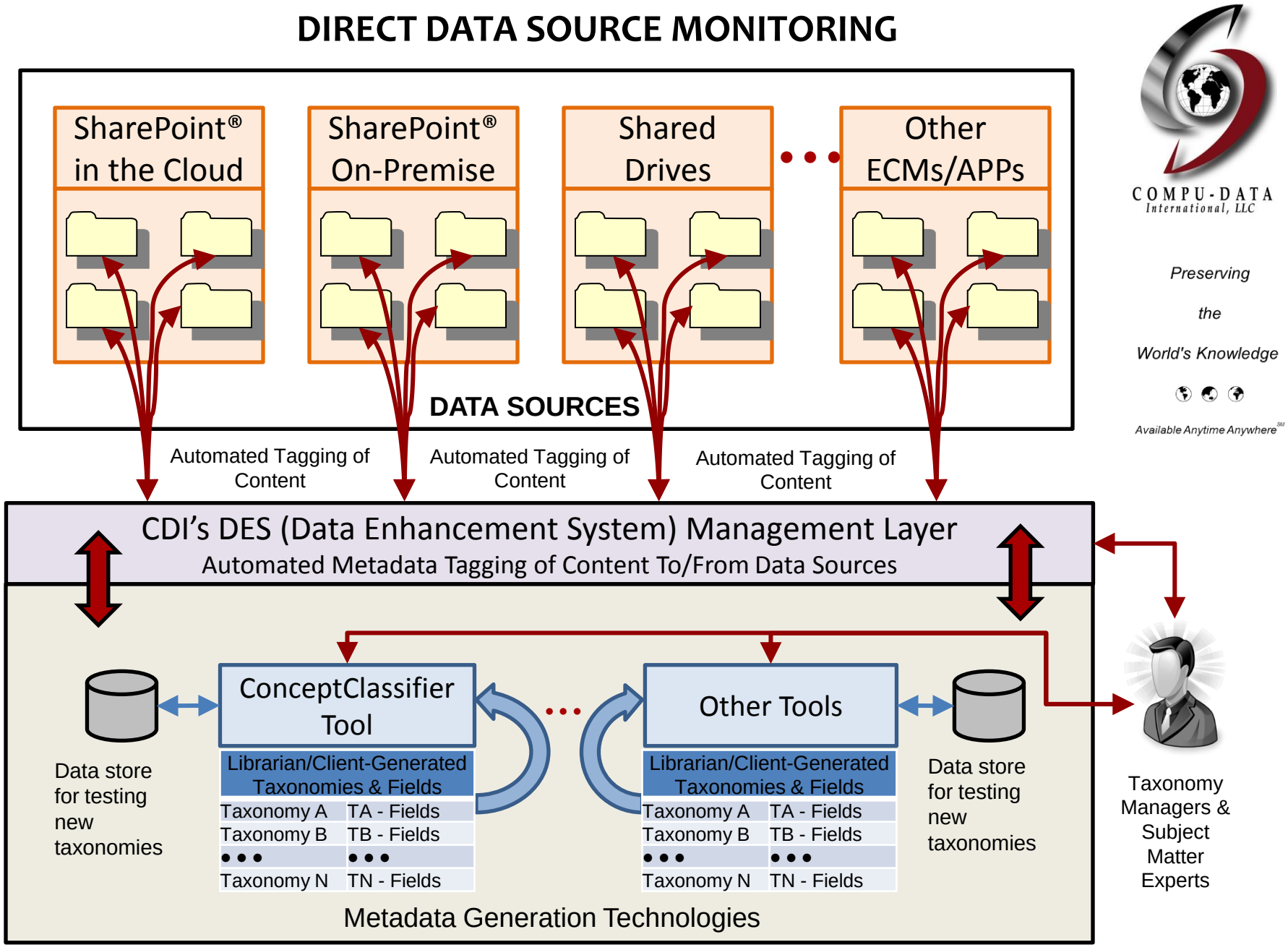
Time Saving Strategies for Applying Managed Metadata to Content Types in a Centralized, Reusable Way


Document Library Settings
Edit Policy
Information Rights Managem...
Page
Safety
Tools
Site Actions
SHAREPOINT\Administrato
Home ▸ Shared Documents ▸ Document Library Settings
I Like It
Home
Libraries
Site Pages
Shared Documents
FOUO
Records Center
Human Resources
ITAR
Drop Off Library
Lists
Calendar
Tasks
Discussions
Team Discussion
Recycle Bin
All Site Content
List Information
Name: Shared Documents
Web Address: http://sp2010srv/Shared Documents/Forms/AllItems.aspx
Description: Share a document with the team by adding it to this document library.
General Settings
Title, description and navigation
Versioning settings
Advanced settings
Validation settings
Column default value settings
Rating settings
Audience targeting settings
Metadata navigation settings
Per-location view settings
Permissions and Management
Delete this document library
Save document library as template
Permissions for this document library
Manage files which have no checked in version
Information Rights Management
Workflow Settings
Generate file plan report
Enterprise Metadata and Keywords Settings
Information management policy settings
Communications
RSS settings
Content Types
This document library is configured to allow multiple content types. Use content types to specify the information you want to display about an item, in addition to its policies, workflows
other behavior. The following content types are currently available in this library:
Content Type
Visible on New Button
Default Content Type
Document
Trusted sites | Protected Mode: Off


Document Library Settings
Edit Policy
Information Rights Managem...
Page
Safety
Tools
Site Actions
SHAREPOINT\Administrato
Home › ... Document Library Settings › Document › Edit Policy
I Like It
Home
Search this site...
Libraries
Site Pages
Shared Documents
FOUO
Records Center
Human Resources
ITAR
Drop Off Library
Lists
Calendar
Tasks
Discussions
Team Discussion
Recycle Bin
All Site Content
Name and Administrative Description
The name and administrative description are shown to list managers when configuring policies on a list or content type.
Name:
Document
Administrative Description:
Declared Records Staging Area
Policy Statement
The policy statement is displayed to end users when they open items subject to this policy. The policy statement can explain which policies apply to the content or indicate any special handling or information that users need to be aware of.
Policy Statement:
Records Staging
Retention
Schedule how content is managed and disposed by specifying a sequence of retention stages. If you specify multiple stages, each stage will occur one after the other in the order they appear on this page.
Note: If the Library and Folder Based Retention feature is active, list administrators can override content type policies with their own retention schedules. To prevent this, deactivate the feature on the site collection.
Enable Retention
Specify how to manage retention:
Items will not expire until a stage is added.
Add a retention stage...
Auditing
Specify the events that should be audited for documents and items subject to this policy.
Enable Auditing
Barcodes
Assigns a barcode to each document or item. Optionally, Microsoft Office applications can require
Enable Barcodes


Document Library Settings
Edit Policy
Information Rights Mana...
Page
Safety
Tool
Site Actions
SHAREPOINT\Administrat
Home › Information Rights Management Settings
I Like It
Home
Libraries
Site Pages
Shared Documents
FOUO
Records Center
Human Resources
ITAR
Drop Off Library
Lists
Calendar
Tasks
Discussions
Team Discussion
Recycle Bin
All Site Content
Information Rights Management (IRM)
IRM helps protect sensitive files from being misused or distributed without permission once they have been downloaded from this server.
Restrict permission to documents in this library on download:
Permission policy title:
FOUO - Sensitive but Unclassified
Example: Company Confidential
Permission policy description:
this stuff is restricted to organization domain only
Example: Only discuss the contents of this document with other employees
Allow users to print documents
Allow users to access content programmatically
Users must verify their credentials every:
2
days
Do not allow users to upload documents that do not support IRM
Stop restricting permission to documents in this library on:
3/31/2011

DIRECT DATA SOURCE MONITORING



COMPU-DATA International, LLC

Preserving the World's Knowledge

Available Anytime Anywhere℠

INTEGRATION WITH CAPTURE & PROCESSES

COMPU-DATA International, LLC

Preserving the World's Knowledge

Available Anytime Anywhere℠

Business Processes

Kodak Capture Pro

Workflow

Info Activate

Automated Tagging & Delivery of Content

SharePoint® In the Cloud

SharePoint® On-Premise

Shared Drives

Others ECMs

DATA SOURCES

CDI's DES (Data Enhancement System) Management Layer

Automated Metadata Tagging of Content To/From Data Sources

Data store for testing new taxonomies

ConceptClassifier Tool

Librarian/Client-Generated Taxonomies & Fields	
Taxonomy A	TA - Fields
Taxonomy B	TB - Fields
• • •	• • •
Taxonomy N	TN - Fields

Other Tools

Librarian/Client-Generated Taxonomies & Fields	
Taxonomy A	TA - Fields
Taxonomy B	TB - Fields
• • •	• • •
Taxonomy N	TN - Fields

Data store for testing new taxonomies

Taxonomy Managers & Subject Matter Experts

Metadata Generation Technologies

Let's Tie all of this Together With a Video Demo

You can see a similar presentation with the video in CDI's YouTube Channel

http://bit.ly/n16bmi

You can see the video I am going to show in CDI's YouTube Channel

http://bit.ly/NHoxkO

COMPU-DATA International, LLC

Cost Effective & Standardized Method to Comply with Records Management & Information Assurance Guidelines

Promote Consistency & Improve Efficiency at Every Level of the Organization

Enforcing Best Practices Around Records Management & Information Assurance

Allowing Document Visibility & Collaboration

Agile & Effective Practices that Enhance Ability to Anticipate, Manage & Respond to Changing Requirements

Enables Organizations to Leverage Metadata & Content Types to Deliver a Faster & More Effective Process of Accessing, Storing, Preserving & Securing Information

Preserving the Worlds Knowledge - Available Anytime Anywhere[SM]

COMPU-DATA International, LLC

Juan J. Celaya
jcelaya@cdlac.com
www.cdlac.com
blog.cdlac.com

431 Nursery Road, Suite A-300, Spring, Texas 77380 (281) 292-1333 x301